

Full Title: Epigenome-wide association study for dilated cardiomyopathy in left ventricular heart tissue identifies putative gene sets associated with cardiac development and early indicators of cardiac risk (short: EWAS of DCM in Heart Tissue)

First author's surname: Tan

Authors' names, academic degrees, and affiliations:

Konstanze Tan (BS)¹, Darwin Tay (PhD)¹, Wilson Tan (PhD)²⁻⁴, Hong Kiat Ng (PhD)¹, Eleanor Wong (PhD)⁴, Michael P Morley (MS)⁵, Gurpreet K Singhera (PhD)^{6,7}, Pritesh R Jain (PhD)¹, Fei Li Tai (BS)¹, Paul J Hanson (PhD)^{7,8}, Thomas P Cappola (MD, ScM)⁵, Kenneth B Margulies (MD)^{5,9-11}, Roger Foo (MD)^{2,3,12-15*}, Marie Loh (PhD)^{1,4,16,17*}

¹ Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

² Cardiovascular Research Institute, National University Health System, Singapore

³ Cardiovascular Disease Translational Research Programme, National University of Singapore, Singapore

⁴ Genome Institute of Singapore (GIS), Agency for Science, Technology and Research (A*STAR), Singapore

⁵ Penn Cardiovascular Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

⁶ Department of Medicine, University of British Columbia, Vancouver, BC, Canada

⁷ Bruce McManus Cardiovascular Biobank, UBC- Centre for Heart Lung Innovation, Vancouver, BC, Canada

⁸ Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

⁹ Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

¹⁰ Department of Physiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

¹¹ Penn Muscle Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

¹² National University Hospital, Singapore

¹³ Yong Loo Lin School of Medicine, National University of Singapore, Singapore

¹⁴ The Clinician Scientist Academy, National University Health System, Singapore

¹⁵ Centre for Translational Medicine, National University Heart Centre, National University Health System, Singapore

¹⁶ Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom

¹⁷ National Skin Centre, Singapore

*These authors jointly supervised the work

Corresponding author details:

Name: Marie Loh

Email: marie_loh@ntu.edu.sg

Complete Address: Lee Kong Chian School of Medicine, Nanyang Technological University, Clinical Sciences Building, 11 Mandalay Road, Singapore 308232.

Telephone: +65 65923076

Total word count of the manuscript (including the title page, abstract, text, references, tables and figures legends): 7,792

ABSTRACT

Background: Methylation changes linked to dilated cardiomyopathy (DCM) affect cardiac gene expression. We investigate DCM mechanisms regulated by CpG methylation using multi-omics and causal analyses in the largest cohort of left ventricular tissues available.

Methods: We mapped DNA methylation at ~850,000 CpG sites, performed array-based genotyping and RNA sequencing in left-ventricular tissue samples from failing and non-failing hearts across two independent DCM cohorts (discovery n=329, replication n=85). Summary data-based Mendelian Randomization (SMR) was applied to explore the causal contribution of sentinel CpGs to DCM. Fine-mapping of regions surrounding sentinel CpGs revealed additional signals for cardiovascular disease risk factors. Coordinated changes across multiple CpG sites were examined using weighted gene correlation network analysis (WGCNA).

Results: We identified 194 epigenome-wide significant CpGs associated with DCM (discovery $P < 5.96E-08$), enriched in active chromatin states in heart tissue. Amongst these, 183 sentinel CpGs significantly influenced the expression of 849 proximal genes ($\pm 1\text{Mb}$). SMR suggested the causal contribution of two sentinel CpGs to DCM and 36 sentinel CpGs to the expression of 43 unique proximal genes ($P < 0.05$). Colocalization analyses indicated that a single causal variant may underlie the methylation-gene expression relationship for three sentinel CpGs. Fine-mapping revealed additional signals linked to cardiovascular traits including hsCRP and blood pressure. Co-methylation modules were enriched in genes related to cardiac

physiological and pathological processes and their corresponding transcriptional regulators.

Conclusions: Using the largest series of left ventricular tissue to date, this study investigates the causal role of cardiac methylation changes in DCM and suggests targets for experimental studies to probe DCM pathogenesis.

INTRODUCTION

Dilated cardiomyopathy (DCM) is the most common form of cardiomyopathy and a leading indication for heart transplantation, affecting 1 in 250-400 individuals, with an incidence of 5-7 cases per 100,000 persons per year.¹ DCM is characterized by a thinning and weakening of the left ventricular heart walls, with resulting contractile dysfunction.²⁻⁴ Clinical diagnostic criteria for DCM include a left ventricular ejection fraction (LVEF) of <45% and a left ventricular end-diastolic dimension >112% of that predicted for age and body surface area.⁵ There is marked heterogeneity in both the prognosis and age of onset for DCM, with most patients becoming symptomatic between 20-50 years of age.⁶ DCM accounts for a substantial 2-3% of yearly sudden cardiac death (SCD) events in which the heart stops beating, resulting in a sudden loss of blood flow to vital organs.⁷

Over the past decade, the genetic basis for DCM has only been partially unraveled.⁸ Beyond genetic predisposition, environmental factors could influence the incidence and course of DCM.⁴ DNA methylation, a key regulator of gene expression, as well as a molecular integrator of genetic and environmental influences, has been investigated in DCM to gain insights into its molecular pathogenesis. To date, multiple epigenome-wide association studies (EWAS) of DCM have identified a handful of differentially methylated loci.⁹⁻¹⁴ This includes the reporting of conservation in methylation patterns between myocardium and peripheral blood at the *NPPA* and *NPPB* loci encoding cardiac stress markers, which are currently gold standard indicators for heart failure prognosis in clinics, thereby demonstrating the potential of DNA methylation as a novel class of DCM biomarkers. However, these epigenome-wide methylation studies of DCM have largely been limited by sample size (n=8-72) owing to the scarcity of left ventricular tissue.

In this study, we sought to expand our understanding of myocardial-specific methylation changes in DCM by analyzing intra-operative left ventricular (LV) tissue from a cohort of 329 individuals of White and African American ancestries (n = 159 cases, 170 controls). This study represents the largest investigation of DCM methylation conducted in myocardial tissue to date. Through integrative omics, causal analyses and fine-mapping of putative methylation markers, we aim to elucidate the putative causal involvement of methylation alterations to DCM pathogenesis.

METHODS

Description of samples

For each cohort, written informed consent for the research use of donated left ventricular (LV) tissue were obtained. For heart transplant recipients, consent was obtained from the transplant recipient. For brain-dead organ donors, consent was obtained from the next-of-kin. All analyses and study protocols were approved by the relevant institutional review boards.

Myocardial Applied Genomics Network Cohort (MAGNet)

LV free-wall tissue was harvested at the time of cardiac surgery from subjects with heart failure undergoing transplantation and from donor hearts deemed unsuitable for transplantation but with apparently normal ventricular function. Hearts were perfused immediately with cold in-situ high-potassium cardioplegia before cardiectomy to arrest contraction and prevent ischemic damage. Tissue samples were promptly frozen in liquid nitrogen. Only samples diagnosed with 'idiopathic DCM' were analysed.

Bruce McManus Cardiovascular Biobank Cohort (BMCB)

LV free-wall tissue (apical portion) was obtained from biopsies of patients with DCM. Biopsies of 1- to 2-mm diameter were washed in ice-cold saline (0.9% NaCl) and stored in liquid nitrogen until DNA extraction. Control tissue samples were non-postmortem donor hearts recovered within 0-2 hours of circulatory death from the International Institute for the Advancement of Medicine (IIAM). Samples with confirmed or suspected non-idiopathic etiology for DCM (e.g., cardiotoxicity, inflammation, familial, myocarditis, ischemia) were excluded.

The iHealth-T2D study

iHealth-T2D is a prospective study of Indian Asian men and women living in West London, and recruited in 2016. At enrolment all participants completed a structured assessment of cardiovascular and metabolic health, including (i) history for CVD such as myocardial infarction, angina, and coronary heart disease; (ii) indicators of CVD risk including blood pressure (systolic and diastolic), Framingham Coronary Heart Disease score, and hypertension/history of hypertension, and (iii) collection of blood samples for measurement of lipid profile and markers such as high-sensitivity C-reactive protein (hsCRP) and creatinine. Methylation profiling was performed using genomic DNA from peripheral blood collected at enrolment. The study is approved by the National Research Ethics Service and all participants gave written informed consent.

Quantification of DNA methylation

Genomic DNA was bisulfite converted using the EZ DNA methylation kit according to the manufacturer's instructions (Zymo Research). Case and control samples were randomized in all experiments. Bisulfite-converted genomic DNA was quantified using the Infinium MethylationEPIC BeadChip (EPIC array). Retrieval of

bead intensity and Illumina background correction was performed using the *minfi* R package (version 1.40.0). Markers were excluded from analyses if they: (i) were non-CpG; (ii) were present on sex chromosomes; or (iii) had low call-rate (<95%). Samples were excluded from analyses if they: (i) had low call-rate (<98%); (ii) had mislabelled gender; or (iii) were present in duplicate. After filtering, the discovery cohort (MAGNet) had 838,624 autosomal CpG probes and 329 samples, while the replication cohort (BMCB) had 841,904 autosomal CpG probes and 85 samples available for analysis.

Statistical analyses of epigenome-wide data

Epigenome-wide methylation data was analysed using the CPACOR pipeline (incorporating Control Probe Adjustment and reduction of global CORrelation) as previously described.¹⁶ Quantile-normalised marker intensity values were then used to calculate per-CpG methylation beta values. Principal component analysis (PCA) was performed on built-in control probe intensities to identify PCs accounting for ~95% of control probe variation which were included as predictors in regression models to adjust for technical biases (MAGNet: first 10 PCs; BMCB: first 3 PCs). The association of each autosomal CpG site with DCM was tested using logistic regression, adjusting for age, gender and control probe PCs (Model: DCM ~ Beta(QN) + age + gender + PCs_{control-probes}). Within the MAGNet cohort, regression was performed separately for each ancestry group, followed by a trans-ancestry inverse variance-weighted meta-analysis using METAL. Test-statistic bias and inflation in association results was adjusted pre- and post- meta-analysis, using a Bayesian algorithm implemented in the *bacon* R package (version 1.30.0).²⁷

Enrichment analysis of genomic regulatory features

Sentinel CpGs were evaluated for enrichment in genomic regulatory features compared to 1000 permutations of EPIC array CpGs. Enrichment was determined through a one-tailed permutation test, computing a P value by comparing overlap counts of the test set (sentinel CpGs) with the distribution from the background. Enrichment in eQTM and transcription factor binding sites (TFBS) was examined against EPIC array CpG sites matched by methylation levels and variability using a sliding-window approach, wherein adjustments were made to mean and standard deviation values until 1000 permuted sites were reached.²⁸ Known TFBS (N=1,210) were obtained from the Remap 2022 database. Enriched TFs were classified as 'Cardiac TF' if expressed in humans and validated *in vivo*^{29,30}, 'Computational-based' if predicted as cardiac TFs computationally due to disruption of TFBS near heart-failure linked loci^{31,32}, or 'non-cardiac TF' if there is no suggestion for a cardiac role in humans in existing literature. eFORGE V2 software assessed sentinel CpG enrichment in learned chromatin states, DNase 1 hotspots, and regions marked by core histone marks across various tissues and cell subsets within the Roadmap Epigenomic Consortium database.³³ For constructing the background set, eFORGE V2 software matched EPIC array CpG sites to sentinel CpGs based on gene and CpG island annotations.

RNA sequencing

Gene expression (RNAseq) data was obtained from the same left ventricular samples analysed for methylation, where available. Prior to downstream analyses, genes were removed if they: (i) had low read counts (<10 counts in the minimum group size of samples) or (ii) were present on sex chromosomes. Gene expression

levels were normalized using TMM (Trimmed Mean of M-values), recommended for quantitative trait loci analyses.³⁴ More details on the sequencing, alignment, filtering and normalization of RNAseq data are available in Supplementary Methods.

Expression quantitative trait methylation (eQTM) analysis

Expression quantitative trait methylation (eQTM) analysis was conducted to identify significant associations between sentinel CpG methylation and proximal gene expression (+/- 1Mb of the sentinel CpG based on gene TSS). Gene expression was modelled against CpG methylation with adjustment for age, gender, ethnicity, RNA Integrity number (RIN) and the top 5 PEER factors which represent latent sources of variability in the gene expression data, possibly including technical and unknown confounders.³⁵ Association testing was performed using a linear model implemented within the *MatrixEQTL* R package (version 2.3).¹⁵

Methylation quantitative trait loci analysis (meQTL)

DNA samples were obtained from the same left ventricular samples analysed for methylation, where available (MAGNet). Genotyping was performed using the Affymetrix Genome-Wide SNP Array 6.0, as previously described.³⁶ SNPs were removed from the current analyses if they: (i) were multiallelic; (ii) were present in duplicate); (iii) had low imputation quality ($R^2 < 0.3$) or (iv) were rare (ancestry-specific $MAF < 0.05$). Methylation quantitative trait loci (meQTL) analysis was conducted to identify significant associations between SNPs and sentinel CpG methylation (+/-500kb). CpG methylation was regressed against SNP dosage values with adjustment for age, gender, methylation array control probe PCs capturing 95% of control probe variation. Association testing was performed using a linear model

implemented within the *MatrixEQTL* R package (version 2.3).¹⁵ Regression was performed separately for each ancestry group, followed by a trans-ancestry inverse variance-weighted meta-analysis using METAL.

Causal analyses (Mendelian Randomisation and Colocalisation)

Summary data-based Mendelian Randomization (SMR) was performed to investigate the putative causal contribution of sentinel CpGs to DCM. SMR uses summary-level association statistics from independent association studies to compute a causal estimate for the influence of an exposure (e.g. sentinel CpG methylation) on an outcome (e.g. DCM disease status, or gene expression).²⁴ The significance of the causal estimate is determined using the Wald Test. Separate SMR analyses were conducted to examine causal relationships between sentinel CpG methylation and: i. DCM ii. proximal gene expression (+/-1Mb). For each sentinel CpG, the lead meQTL ($P < 0.05$) that was also analysed for association with the respective outcome was chosen to be the instrumental variable (IV) for assessing causal relationships. Genetic associations for DCM and gene expression were obtained from the largest published GWAS of DCM ($n = 355,381$, UKBioBank) and left ventricular cis-eQTL study ($n = 386$; GTEx v8 release; SNPs within +/- 1Mb of gene transcription start sites), respectively. We validated SMR-significant associations using one-sample MR and conducted colocalisation analyses to evaluate the posterior probabilities of a shared causal variant for the assessed traits. One-sample Mendelian Randomization (MR) was performed with individual-level genotype, methylation, and outcome data (MAGNet) using the 2-stage least squares regression method implemented in the *AER* R package (version: 1.2-12). Colocalization analysis was conducted using the *coloc.abf* function in the *coloc* R

package (version 5.2.3), based on a +/-500kb region around each sentinel CpG as this was the window size used for identifying meQTL.³⁷

Weighted gene correlation network analysis (WGCNA)

We employed Weighted Gene Correlation Network Analysis (WGCNA) to construct co-methylation modules using DCM-associated CpGs from discovery-stage EWAS.³⁸ These CpGs were (i) associated with DCM (FDR $P < 0.05$); (ii) replicated with consistent association direction; and (iii) displayed variability (methylation beta SD > 0.02). Using 32,198 CpGs we built a signed consensus co-methylation network using the *blockwiseconsensusModules* function in the *WGCNA* R package (version 1.72-5), utilizing settings recommended for our sample size (soft thresholding power = 12, merge cut height = 0.25 and minimum module size = 30). Methylation levels within consensus modules were summarized as module eigengenes (ME) and tested for correlation with DCM separately by ancestry (Model: $DCM \sim ME + Age + Gender + 10PCs$). Inverse variance-weighted meta-analysis combined ancestry-specific results to determine module associations with DCM. Additionally, a hypergeometric test was applied to assess module enrichment in DCM sentinel CpGs ($P < 0.05$) against all CpGs used for network construction. Genes mapped to CpGs within co-methylation modules for overrepresentation of Gene Ontology (GO) terms and biological processes from KEGG and REACTOME databases using the *clusterProfiler* R package (version 4.10.0). GO gene sets were obtained from *org.Hs.eg.db* (version 3.18.0), while KEGG and REACTOME datasets were accessed via the *msigdb* R package (version 7.5.1).

Targeted methylation sequencing

Targeted methylation sequencing was performed using genomic DNA from peripheral blood samples collected from participants of the iHealth-T2D study. Targeted methylation sequencing was performed for regions defined as +/-500bp of DCM sentinel CpGs, supported by previous publications demonstrating a decay in correlation between methylation sites beyond 1-2kb,^{39,40} and from our in-house whole-genome bisulfite sequencing dataset where we observed that $|r|>0.2$ was within +/-500bp for most of the assessed CpGs (data unpublished). Additional details on probe design, sample processing and library preparation, sequencing and quality control steps employed are available in Supplementary Methods.

Statistical analyses of targeted methylation sequencing data

Within each sequenced region, the association of each CpG site with CVD-related traits was tested using linear regression for continuous traits and logistic regression for binary traits, adjusting for age, gender and white blood cell proportion of six white blood cell sub-populations (CD8 T cells, CD4 T cells, Natural Killer Cells, B-cells, Monocytes, Granulocytes) estimated by the Houseman algorithm.⁴¹ In parallel, pairwise correlation in methylation levels between CpGs within regions was calculated using the R function *cor*, which computes the correlation coefficient. For a given pair of CpGs, only samples where methylation data was available for both CpGs being compared were used to calculate correlation.

Construction of methylation score

We constructed a methylation score from CpGs in fine-mapped regions and examined the association of these scores with CVD traits. Effect sizes of association between individual CpG loci and CVD traits were used as weights for constructing

the methylation score. Specifically, we defined trait-specific methylation scores within each region as a linear combination of k CpG site beta values b and weights w :

$$\text{MRS} = \sum_{i=1}^k w_i b_i$$

Within each region, the association of each MRS with CVD-related traits was assessed, adjusting for the same covariates as in single loci association testing described in '***Statistical analyses of targeted methylation sequencing data***'.

RESULTS

Overview of study design

Our study design is summarised in Figure 1. In brief, we first carried out an epigenome-wide association investigation of DCM using 414 left ventricular samples obtained from two repositories: the Myocardial Applied Genomics network (MAGNet; $n=329$ [discovery]) and Bruce McManus Cardiovascular Biobank (BMCB; $n=85$ [replication]) (Supplementary Table 1). Integrative omics analyses were performed on the identified sentinel CpGs. To discover additional signals beyond CpG sites captured by the methylation array, we conducted fine-mapping of selected top-ranking loci in blood samples obtained from a population-based cohort (iHealth-T2D, $n=1974$).

Epigenome-wide association analysis

We performed EWAS of DCM using genomic DNA extracted from LV free-wall tissue, with separate EWAS carried out for Whites and African Americans within the discovery cohort (MAGNet), followed by inverse-variance meta-analysis (Methods). From discovery-stage EWAS, 196 CpG sites were associated with DCM at a

Bonferroni-corrected threshold of $P < 5.96E-08$ ($0.05/838,624$) (Figure 2). Subsequent targeted replication testing in the BMCB cohort confirmed consistent directionality of effect size estimates for all discovery CpGs (Supplementary Table 2). The 196 CpG sites were distributed across 171 genetic loci, with 150/171 (88%) genetic loci containing a single sentinel CpG and 21/171 (12%) loci containing two or more sentinel CpGs (Supplementary Figure 1, Supplementary Table 2). Conditional analyses at each locus identified a total of 194 robustly associated and conditionally independent signals ('sentinel CpGs'), which were further analysed for functional relevance and causal contribution to DCM pathogenesis.

Unsupervised hierarchical clustering based on the methylation levels of the 194 sentinel CpGs resulted in two distinct clusters, segregating samples by their respective case and control status. The 'case' cluster correctly consisted of 128 (88%) cases, while the 'control' cluster had 153 (83%) controls (binomial $P < 2.2E-16$) (Figure 3). This finding supports a perturbation of DNA methylation in DCM. Clustering by case and control status persisted in ancestry-specific unsupervised hierarchical clustering of methylation levels of the sentinel CpGs (Supplementary Figure 2).

Enrichment of Sentinel CpGs in active gene regulatory regions and impact on proximal gene expression

To understand the regulatory role of sentinel CpGs, we first examined the enrichment of annotated chromatin states. Sentinel CpGs were enriched in transcriptionally active chromatin states and depleted in repressed Polycomb states of LV tissue (permutation test $P < 0.001$), compared to a background of CpGs

matched by relationships to gene and CpG islands (Supplementary Figure 3A). Consistent with their enrichment in active LV chromatin states, sentinel CpGs were also enriched for location in H3K4me3-marked regions, typically representing transcriptionally active chromatin and promoters (permutation test $P < 0.001$) (Supplementary Figure 3B).

To identify sentinel CpGs that impacted proximal gene expression, hereon referred to as expression quantitative methylation loci (eQTM), we examined the association between methylation levels of the 194 sentinel CpGs and expression of their proximal genes (<1Mb) present in the discovery (MAGNet) and replication (BMCB) RNAseq datasets. We found the 194 sentinel CpGs to be enriched for association with proximal gene expression in LV tissue (3.80-fold compared to expectations under the null hypothesis; $P < 0.001$) (Supplementary Figure 4). Subsequent targeted replication testing on sentinel CpG-gene pairs reaching FDR $P < 0.05$ in discovery-stage association testing confirmed consistent directionality of effect size estimates between 183 sentinel CpGs and 849 unique proximal genes ('replicated eQTMs') (Supplementary Table 3).

Summary data-based Mendelian Randomisation to infer disease causation

Having obtained initial evidence of sentinel CpG contribution to transcriptional regulation, we next investigated whether these CpGs had a causal relationship with both DCM and proximal gene expression using Summary data-based Mendelian Randomization (SMR). Separate causal analyses were conducted for DCM and gene expression. We found two sentinel CpGs that were causally linked to DCM (cg08140459 and cg12359658; $p < 0.05$), with subsequent validation testing via one-

sample MR confirming consistent directionality of causal estimate for cg08140459-DCM ($p < 0.05$) (Supplementary Table 4).

To gain further insight into the molecular mechanisms underpinning the contribution of sentinel CpGs to DCM, we conducted a separate SMR of gene expression, focusing on replicated eQTMs. After excluding sentinel CpGs without suitable instrumental variables (e.g., meQTL not included in public (GTEx) dataset), 931 eQTMs (181 unique sentinel CpGs, 828 unique genes) could be analysed in the SMR of gene expression. Out of the 181 sentinel CpGs analysed, 36 sentinel CpGs showed putative causal relationships with 43 unique proximal genes ($p < 0.05$) (Supplementary Table 5).

We further integrated evidence from multi-omics analyses to identify the most relevant sentinel CpGs for DCM pathogenesis. Among the 36 sentinel CpGs causally linked to proximal gene expression, we selected the three CpGs with the highest posterior probability for a shared causal variant influencing both CpG methylation and proximal gene expression (cg09862509-*IER5*, coloc.abf-PP.H4 = 0.91; cg11793257-*ENTPD6*, coloc.abf-PP.H4 = 0.69; cg11793257-*ABHD12*, coloc.abf-PP.H4 = 0.47; cg06807905-*KCNC4*, coloc.abf-PP.H4 = 0.51) (Figure 4, Supplementary Figure 5). Causal estimates for these CpGs were validated in a one-sample MR setting, showing consistent direction of association (Supplementary Data 5).

Although the cardiac roles of the target genes have not been well-studied, existing investigations suggest functions with potential relevance to DCM pathogenesis. *IER5* is a transcription factor regulating cell proliferation, possibly via the crosstalk between Notch and DNA damage response pathways. *ENTPD6* is a

nucleotide-metabolising enzyme that is predominantly expressed in the heart and functions in platelet recruitment and aggregation.¹⁷ The potassium ion channel *KCNK4* influences neural cell survival and apoptosis under oxidative stress conditions in mice and may also play a role in regulating myocardial action potential.^{18,19} *ABDH12* is an enzyme that hydrolyses endocannabinoids, a class of lipids regulating a wide range of pathologies including platelet aggregation, vasodilation, and the maintenance of energy balance.²⁰

Genes mapped to CpG sites demonstrating coordinated changes in methylation patterns are enriched in disease-relevant pathways

Beyond analysing single CpG associations, examining coordinated methylation changes across multiple CpG sites and their linked genes could reveal disease-relevant pathways regulated by DCM methylation. To achieve this, we conducted weighted correlation network analysis (WGCNA), constructing co-methylation modules using methylation levels of 32,918 DCM-associated CpG sites (discover EWAS FDR $P < 0.05$, with consistent direction of association in replication EWAS, $SD > 0.02$ across all samples). Seven co-methylation modules were identified (Figure 5). Regressing the module eigengene (the first principal component of module-specific methylation intensities) against DCM confirmed an association with DCM for all modules (Supplementary Table 6).

To investigate the biological relevance of co-methylation modules, we looked for over-represented gene sets and enrichment in binding sites of known transcription factors (TFs). (Figure 5, Supplementary Figure 6) Gene set overrepresentation was analyzed using genes belonging to replicated eQTM pairs of module-specific CpGs.

Multiple modules were enriched in gene sets associated with cardiac development and key pathological processes in heart failure, such as inflammation and alterations in extracellular matrix (ECM) structure and composition (Figure 5, Supplementary Table 7). Notably, Module 5 CpG loci were enriched in binding sites of *YAP*, a key effector of the Hippo pathway that has a well-established and essential role in cardiac development²¹ (Supplementary Figure 6). Certain modules showed enrichment for gene sets and transcription regulators more specific to cardiomyopathy. For instance, Module 1 was enriched for gene sets related to muscle adaptation, including processes such as hypertrophy or remodeling in response to injury (Figure 5). Correspondingly, Module 1 CpG loci were enriched in binding sites of TFs regulating cardiomyocyte hypertrophy, such as *THRA*, *PTRF*, *NSD2*, *HOXA9*, and *LZBTB* ($P < 0.001$) (Supplementary Figure 6). Module 6 was enriched for gene sets related to nucleotide and lipid metabolism, regulation of calcium ion transport and homeostasis, as well as enzymatic activity catalyzing chemical modification of nucleic acids and other biomolecules (Figure 5). Supporting the enrichment of metabolism-related gene sets, binding sites of *RXR*, a transcription factor known to regulate cardiac metabolism, was enriched amongst Module 6 CpG loci ($P < 0.001$)²³ (Supplementary Figure 6).

Fine-mapping sentinel CpGs to investigate regional associations with traits related to cardiac disease and disease risk

As the methylation array covers only 2-3% of CpG sites in the epigenome, we sought to improve our investigation of regional associations using an existing target methylation sequencing dataset to increase coverage of CpG sites (\pm 500bp) surrounding the top-performing DCM sentinel CpGs. Targeted methylation sequencing was performed using blood samples of individuals from the iHealth-T2D

study, which is well phenotyped for various traits relevant to CVD (n=1974). We assessed regional associations with (i) CVD such as myocardial infarction, angina, and coronary heart disease; (ii) indicators of CVD risk including blood pressure (systolic and diastolic), Framingham Coronary Heart Disease score, and hypertension/history of hypertension; (iii) high-sensitivity C-reactive protein (hsCRP), an inflammatory marker and predictor of severe cardiac outcomes in individuals without prior cardiac disease, and (iv) creatinine, a renal marker reflecting renal function that can worsen in heart failure.

Targeted sequencing was performed on regions surrounding 28 DCM sentinel CpGs (28 regions) (Supplementary Table 8). A total of 293 CpG sites were captured, with two to 24 CpG sites captured in each region. Seventeen regions had significant associations with at least one of the 10 unique CVD traits (Bonferroni-corrected $P < 0.05$) (Supplementary Table 9). At 14 regions, the CpGs with the strongest association with our CVD traits were not CpGs on the EPIC array, further illustrating the added value brought upon by targeted sequencing. Multiple independent signals were also found for two regions, whereby conditioning on the lead signal revealed secondary signals in both regions.

To illustrate the utility of targeted sequencing to improve regional associations with DCM, we highlight regions surrounding two DCM sentinel CpGs that contained signals for CVD-relevant traits (Bonferroni-corrected $P < 0.05$) from CpGs not present on the EPIC array (Figure 6A-C). In the region surrounding cg13173392 (EWAS $P = 4.71E-09$; third most significant hit), targeted methylation sequencing identified stronger CHD and DBP signals from CpGs that were correlated with cg13173392 (chr11_69245749 with CHD, $P = 1.93E-03$, $|r| = 0.46$; chr11_69245240 with DBP,

$P=1.29E-03$, $|r| = 0.24$) (Figure 6A,B), while cg13173392 itself was only nominally significant for both CHD ($P=2.26E-01$) and DBP ($P=1.53E-01$).

In a separate region surrounding cg16429725 (EWAS $P=5.33 E-09$, sixth most significant hit), a CpG correlated with cg16429725 was identified as a strong hsCRP signal (chr16_57798306, $P=2.79E-05$, $|r| = 0.57$; Figure 6C), while cg16429725 itself was only nominally significant for hsCRP ($P=2.16E-02$). The investigated region surrounding cg16429725 is within the *KIFC3* gene involved in the formation, maintenance, and remodeling of the bipolar mitotic spindle. Although *KIFC3* has not been studied in a cardiac context, its potential contribution to inflammation is supported by its contribution to the hepatocellular carcinoma progression via the PI3K/AKT pathway and its role in stabilizing E-cadherin.^{24,25}

We next combined methylation information from multiple CpGs in each region using a weighted methylation risk score to investigate associations with CVD. Of the 28 sequenced regions, 25 regions had significant methylation risk scores (Bonferroni $P<0.05$) for at least one investigated CVD trait. In 23 of these 25 regions, individual CpGs were not significantly associated with CVD (Bonferroni $P<0.05$). However, combining the methylation patterns of multiple CpGs within these regions into an MRS revealed significant associations with at least one of the investigated CVD traits, indicating that the combined effect of multiple CpGs provides stronger association with CVD traits than individual CpGs alone (Supplementary Table 10).

DISCUSSION

We perform the largest EWAS of DCM in cardiac tissues to date (discovery $n = 159$ DCM, 170 control), extending on previous EWAS of DCM in terms of sample size and coverage of CpG sites. We identified and replicated 194 independent CpG

signals for DCM that reached epigenome-wide significance ($P < 5.96E-08$), with comprehensive multi-omics and causal analyses suggesting the causal contribution of a subset of our sentinels to DCM pathogenesis and transcriptional regulation. Fine-mapping of putative methylation markers and network analysis of coordinated changes across multiple DCM-associated CpG loci supported the relevance of DCM-linked methylation changes to cardiac development, disease pathogenesis as well as early indicators of cardiac risk.

Before the current investigation, the largest existing EWAS of DCM was conducted in a predominantly White cohort (discovery $n=41$ DCM, 31 control) using the older 450k array. This study used a staged multi-omics approach to identify CpG sites with conserved DCM-linked methylation alterations and impact on proximal gene expression.⁹ Among these, cg24884140 stood out as a promising diagnostic biomarker, showing consistent dysmethylation in both heart and blood, a strong link to gene expression, and outperforming the clinical standard heart failure biomarker N-terminal pro-BNP in ROC analysis for detecting DCM. Although cg24884140 did not qualify as one of the 194 sentinel CpGs in the current investigation, it was associated with DCM in the discovery-stage EWAS ($\beta = -12.67$, $P = 8.70E-04$), with the directionality of the association confirmed in subsequent replication testing ($\beta = -53.10$, $P = 1.19E-02$).

With individual-level genotype and transcriptome data available in addition to DNA methylation data, we could conduct multi-omics analyses that supported causal involvement of certain sentinel CpGs in DCM pathogenesis. The 194 sentinel CpGs were enriched in active chromatin states in heart tissue and significantly correlated with proximal gene expression more than expected by chance, supporting a role in transcriptional regulation. These preliminary indications of disease relevance probed

us to investigate the putative causal contribution of sentinel CpGs to DCM. While individual-level genotype data enables the application of one-sample Mendelian Randomization to address population heterogeneity, Summary data-based Mendelian Randomisation (SMR) leverages data from large-scale association studies with higher statistical power to derive causal estimates. Therefore, we opted to conduct SMR as the primary analysis and one-sample MR to validate SMR causal estimates reaching nominal significance ($P < 0.05$). We show putative causal relationships between cg08140459 and DCM. In our study, cg08140459 was robustly linked to the expression of *LTBP2*, a recently discovered prognostic biomarker for DCM, in independent cohorts.²⁶ To gain insight into the molecular pathways in DCM pathogenesis that involve sentinel CpGs, we conducted a separate SMR analysis for gene expression, revealing the putative contribution of three sentinels to nearby genes (cg09862509-*IER5*, cg11793257-*ENTPD6*, cg11793257-*ABHD12*, cg06807905-*KCNC4*) with functions relevant to DCM pathogenesis, thus warranting further investigation in a cardiac context.

Using targeted methylation sequencing data from a population-based cohort phenotyped for multiple cardiac traits, we discovered independent CVD risk factor signals near DCM sentinels that were not captured by the EPIC array. While individual CpGs might demonstrate modest associations with CVD traits, combining methylation data from CpGs on a regional level strengthened these associations. This illustrates the value of targeted methylation sequencing in uncovering loci relevant to CVD traits.

Expanding our analysis beyond individual CpG associations, we employed co-methylation network analysis using CpGs associated with DCM in our discovery cohort (FDR $P < 0.05$) to uncover pathways affected by methylation alterations in

DCM. While gene sets associated with general cardiac pathologies such as inflammation and ECM alterations were enriched across multiple modules, gene sets related to muscle adaptation, metabolism, and calcium ion transport, which are more closely linked to cardiomyopathy were only enriched in specific modules. These findings indicate that CpG loci associated with DCM in our study regulate broad cardiac pathologies as well as those specific to cardiomyopathy

Our study has some limitations. Firstly, the cell type heterogeneity of left ventricular tissue makes it challenging to delineate the specific cell types driving the associations between CpG methylation and DCM. Although bioinformatics cell type deconvolution methods exist for this task, the lack of reference methylation profiles for heart cell types means that only reference-free approaches can be applied. While reference-free deconvolution algorithms can predict distinct cell classes using major variations in methylation profiles, finding a biological basis to justify assigning these output classes to specific heart cell types (e.g. cardiomyocytes or cardiofibroblasts) currently poses a significant challenge. As single-cell profiling techniques for methylation and gene expression in cardiac cell types advance, future methylation studies of DCM should prioritise elucidating the cell type specificity of DCM-linked CpG methylation and the genes they regulate. Concerning our MR analyses, one limitation would be the potential bias in the causal relationships estimated by two-sample MR owing to population heterogeneity between the multi-ancestry MAGNet cohort (White, African American) used to generate meQTL and the predominantly White cohorts in which the GWAS of DCM and left ventricular eQTL analyses were conducted. Despite this, we did not restrict the meQTL analysis to only samples of White ancestry in our cohort with methylation and genotype data available (n=194) to maximise power for detecting left ventricular meQTL associations. On a related note,

a second limitation of the MR analyses would be the limited sample size of the cohort used to generate meQTL, despite the current meQTL analysis being the largest to be conducted in left ventricular tissue to date.

Nonetheless, our study has important strengths. Besides being the largest existing study of DCM-linked methylation disturbances in left ventricular tissue to date, the current investigation extends previous methylation studies of DCM by seeking evidence for the causal contribution of DCM-linked sentinel CpGs and by investigating regional associations with traits indicative of cardiac risk.

CONCLUSION

This is the largest investigation of perturbed CpG methylation in DCM to be conducted in disease-relevant left ventricular tissue obtained from patients and controls. We identify CpGs independently and robustly associated with DCM and suggest molecular players in new, putative causal mechanisms by which DNA methylation may impact DCM. We also provide preliminary indication of the prognostic potential of DCM-linked CpGs associated with cardiac-relevant traits preceding overt disease in healthy individuals.

ACKNOWLEDGEMENTS AND FUNDING

This research is supported by a Start-Up Grant (awarded to Marie Loh [PI]) from Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.

DISCLOSURES

None.

REFERENCES

1. Reichart D, Magnussen C, Zeller T, Blankenberg S. Dilated cardiomyopathy: from epidemiologic to genetic phenotypes: A translational review of current literature. *J Intern Med*. 2019;286:362-372. doi: 10.1111/joim.12944
2. Cuenca S, Ruiz-Cano MJ, Gimeno-Blanes JR, Jurado A, Salas C, Gomez-Diaz I, Padron-Barthe L, Grillo JJ, Vilches C, Segovia J, et al. Genetic basis of familial dilated cardiomyopathy patients undergoing heart transplantation. *J Heart Lung Transplant*. 2016;35:625-635. doi: 10.1016/j.healun.2015.12.014
3. Alraies MC, Eckman P. Adult heart transplant: indications and outcomes. *J Thorac Dis*. 2014;6:1120-1128. doi: 10.3978/j.issn.2072-1439.2014.06.44
4. Schultheiss HP, Fairweather D, Caforio ALP, Escher F, Hershberger RE, Lipshultz SE, Liu PP, Matsumori A, Mazzanti A, McMurray J, et al. Dilated cardiomyopathy. *Nat Rev Dis Primers*. 2019;5:32. doi: 10.1038/s41572-019-0084-1

5. Mathew T, Williams L, Navaratnam G, Rana B, Wheeler R, Collins K, Harkness A, Jones R, Knight D, O'Gallagher K, et al. Diagnosis and assessment of dilated cardiomyopathy: a guideline protocol from the British Society of Echocardiography. *Echo Res Pract*. 2017;4:G1-G13. doi: 10.1530/ERP-16-0037
6. Towbin JA, Lowe AM, Colan SD, Sleeper LA, Orav EJ, Clunie S, Messere J, Cox GF, Lurie PR, Hsu D, et al. Incidence, causes, and outcomes of dilated cardiomyopathy in children. *JAMA*. 2006;296:1867-1876. doi: 10.1001/jama.296.15.1867
7. Halliday BP, Cleland JGF, Goldberger JJ, Prasad SK. Personalizing Risk Stratification for Sudden Death in Dilated Cardiomyopathy: The Past, Present, and Future. *Circulation*. 2017;136:215-231. doi: 10.1161/CIRCULATIONAHA.116.027134
8. Garnier S, Harakalova M, Weiss S, Mokry M, Regitz-Zagrosek V, Hengstenberg C, Cappola TP, Isnard R, Arbustini E, Cook SA, et al. Genome-wide association analysis in dilated cardiomyopathy reveals two new players in systolic heart failure on chromosomes 3p25.1 and 22q11.23. *Eur Heart J*. 2021;42:2000-2011. doi: 10.1093/eurheartj/ehab030
9. Meder B, Haas J, Sedaghat-Hamedani F, Kayvanpour E, Frese K, Lai A, Nietsch R, Scheiner C, Mester S, Bordalo DM, et al. Epigenome-Wide Association Study Identifies Cardiac Gene Patterning and a Novel Class of Biomarkers for Heart Failure. *Circulation*. 2017;136:1528-1544. doi: 10.1161/CIRCULATIONAHA.117.027355
10. Haas J, Frese KS, Park YJ, Keller A, Vogel B, Lindroth AM, Weichenhan D, Franke J, Fischer S, Bauer A, et al. Alterations in cardiac DNA methylation in

- human dilated cardiomyopathy. *EMBO Mol Med*. 2013;5:413-429. doi: 10.1002/emmm.201201553
11. Koczor CA, Lee EK, Torres RA, Boyd A, Vega JD, Uppal K, Yuan F, Fields EJ, Samarel AM, Lewis W. Detection of differentially methylated gene promoters in failing and nonfailing human left ventricle myocardium using computation analysis. *Physiol Genomics*. 2013;45:597-605. doi: 10.1152/physiolgenomics.00013.2013
 12. Jo BS, Koh IU, Bae JB, Yu HY, Jeon ES, Lee HY, Kim JJ, Choi M, Choi SS. Methylome analysis reveals alterations in DNA methylation in the regulatory regions of left ventricle development genes in human dilated cardiomyopathy. *Genomics*. 2016;108:84-92. doi: 10.1016/j.ygeno.2016.07.001
 13. Movassagh M, Choy MK, Goddard M, Bennett MR, Down TA, Foo RS. Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. *PLoS One*. 2010;5:e8564. doi: 10.1371/journal.pone.0008564
 14. Movassagh M, Choy MK, Knowles DA, Cordeddu L, Haider S, Down T, Siggins L, Vujic A, Simeoni I, Penkett C, et al. Distinct epigenomic features in end-stage failing human hearts. *Circulation*. 2011;124:2411-2422. doi: 10.1161/CIRCULATIONAHA.111.040071
 15. Kasturiratne A, Khawaja KI, Ahmad S, Siddiqui S, Shahzad K, Athauda LK, Jayawardena R, Mahmood S, Muilwijk M, Batool T, et al. The iHealth-T2D study, prevention of type 2 diabetes amongst South Asians with central obesity and prediabetes: study protocol for a randomised controlled trial. *Trials*. 2021;22:928. doi: 10.1186/s13063-021-05803-7

16. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Scott J, Jarvelin MR, Elliott P, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.* 2015;16:37. doi: 10.1186/s13059-015-0600-x
17. Yeung G, Mulero JJ, McGowan DW, Bajwa SS, Ford JE. CD39L2, a gene encoding a human nucleoside diphosphatase, predominantly expressed in the heart. *Biochemistry.* 2000;39:12916-12923. doi: 10.1021/bi000959z
18. Song MS, Ryu PD, Lee SY. Kv3.4 is modulated by HIF-1alpha to protect SH-SY5Y cells against oxidative stress-induced neural cell death. *Sci Rep.* 2017;7:2075. doi: 10.1038/s41598-017-02129-w
19. Ordog B, Brutyo E, Puskas LG, Papp JG, Varro A, Szabad J, Boldogkoi Z. Gene expression profiling of human cardiac potassium and sodium channels. *Int J Cardiol.* 2006;111:386-393. doi: 10.1016/j.ijcard.2005.07.063
20. Alfulajj N, Meiners F, Michalek J, Small-Howard AL, Turner HC, Stokes AJ. Cannabinoids, the Heart of the Matter. *J Am Heart Assoc.* 2018;7. doi: 10.1161/JAHA.118.009099
21. Neiningner AC, Dai X, Liu Q, Burnette DT. The Hippo pathway regulates density-dependent proliferation of iPSC-derived cardiac myocytes. *Sci Rep.* 2021;11:17759. doi: 10.1038/s41598-021-97133-6
22. Liu R, Lee J, Kim BS, Wang Q, Buxton SK, Balasubramanyam N, Kim JJ, Dong J, Zhang A, Li S, et al. Tead1 is required for maintaining adult cardiomyocyte function, and its loss results in lethal dilated cardiomyopathy. *JCI Insight.* 2017;2. doi: 10.1172/jci.insight.93343

23. Shao M, Lu L, Wang Q, Ma L, Tian X, Li C, Li C, Guo D, Wang Q, Wang W, et al. The multi-faceted role of retinoid X receptor in cardiovascular diseases. *Biomed Pharmacother.* 2021;137:111264. doi: 10.1016/j.biopha.2021.111264
24. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48:481-487. doi: 10.1038/ng.3538
25. Sako-Kubota K, Tanaka N, Nagae S, Meng W, Takeichi M. Minus end-directed motor KIFC3 suppresses E-cadherin degradation by recruiting USP47 to adherens junctions. *Mol Biol Cell.* 2014;25:3851-3860. doi: 10.1091/mbc.E14-07-1245
26. Nishiura K, Yokokawa T, Misaka T, Ichimura S, Tomita Y, Miura S, Shimizu T, Sato T, Kaneshiro T, Oikawa M, et al. Prognostic Role of Circulating LTBP-2 in Patients With Dilated Cardiomyopathy: A Novel Biomarker Reflecting Extracellular Matrix LTBP-2 Accumulation. *Can J Cardiol.* 2023;39:1436-1445. doi: 10.1016/j.cjca.2023.05.015
27. van Iterson M, van Zwet EW, Consortium B, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* 2017;18:19. doi: 10.1186/s13059-016-1131-9
28. Flam E, Jang C, Murashige D, Yang Y, Morley MP, Jung S, Kantner DS, Pepper H, Bedi KC, Jr., Brandimarto J, et al. Integrated landscape of cardiac metabolism in end-stage human nonischemic dilated cardiomyopathy. *Nat Cardiovasc Res.* 2022;1:817-829. doi: 10.1038/s44161-022-00117-6

29. Yang J, Wang D, Yang Y, Yang W, Jin W, Niu X, Gong J. A systematic comparison of normalization methods for eQTL analysis. *Brief Bioinform.* 2021;22. doi: 10.1093/bib/bbab193
30. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc.* 2012;7:500-507. doi: 10.1038/nprot.2011.457
31. McAllan L, Baranasic D, Villicana S, Brown S, Zhang W, Lehne B, Adamo M, Jenkinson A, Elkalaawy M, Mohammadi B, et al. Integrative genomic analyses in adipocytes implicate DNA methylation in human obesity and diabetes. *Nat Commun.* 2023;14:2784. doi: 10.1038/s41467-023-38439-z
32. Hu R, Morley MP, Brandimarto J, Tucker NR, Parsons VA, Zhao SD, Meder B, Katus HA, Ruhle F, Stoll M, et al. Genetic Reduction in Left Ventricular Protein Kinase C-alpha and Adverse Ventricular Remodeling in Human Subjects. *Circ Genom Precis Med.* 2018;11:e001901. doi: 10.1161/CIRCGEN.117.001901
33. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10:e1004383. doi: 10.1371/journal.pgen.1004383
34. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559. doi: 10.1186/1471-2105-9-559

35. Akazawa H, Komuro I. Roles of cardiac transcription factors in cardiac hypertrophy. *Circ Res*. 2003;92:1079-1088. doi: 10.1161/01.RES.0000072977.86706.23
36. Meganathan K, Sotiriadou I, Natarajan K, Hescheler J, Sachinidis A. Signaling molecules, transcription growth factors and other regulators revealed from in-vivo and in-vitro models for the regulation of cardiac development. *Int J Cardiol*. 2015;183:117-128. doi: 10.1016/j.ijcard.2015.01.049
37. Hannenhalli S, Putt ME, Gilmore JM, Wang J, Parmacek MS, Epstein JA, Morrisey EE, Margulies KB, Cappola TP. Transcriptional genomics associates FOX transcription factors with human heart failure. *Circulation*. 2006;114:1269-1276. doi: 10.1161/CIRCULATIONAHA.106.632430
38. Tan WLW, Anene-Nzelu CG, Wong E, Lee CJM, Tan HS, Tang SJ, Perrin A, Wu KX, Zheng W, Ashburn RJ, et al. Epigenomes of Human Hearts Reveal New Genetic Variants Relevant for Cardiac Disease and Phenotype. *Circ Res*. 2020;127:761-777. doi: 10.1161/CIRCRESAHA.120.317254
39. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol*. 2011;12:R10. doi: 10.1186/gb-2011-12-1-r10
40. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*. 2006;38:1378-1385. doi: 10.1038/ng1909
41. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate

measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86. doi:
10.1186/1471-2105-13-86

FIGURES WITH FIGURE LEGENDS

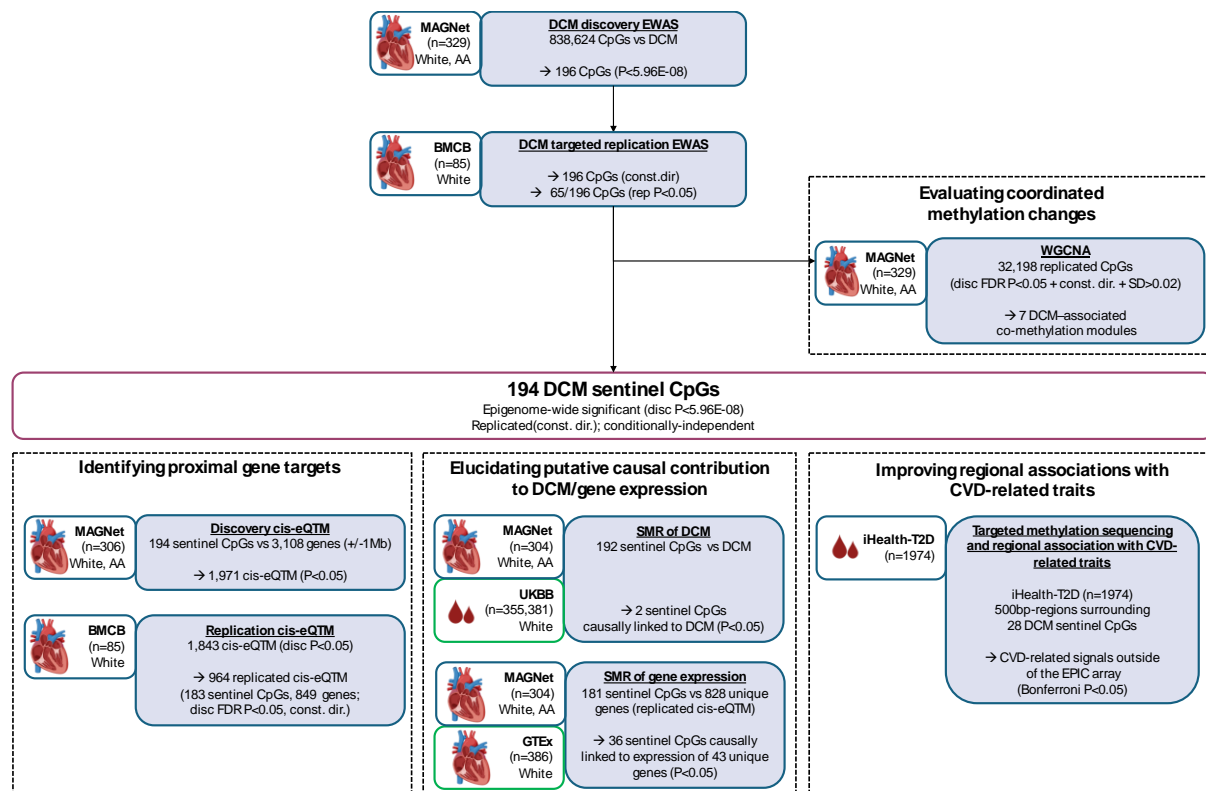


Figure 1 Study Design. Key abbreviations used: (Analyses) EWAS=epigenome-wide association study; eQTM=expression quantitative trait methylation analysis; (Ancestries) AA=African American; (Cohorts) MAGNet=Myocardial Applied Genomics Network; BMCB=Bruce McManus Cardiovascular Biobank; UKBB=UK Biobank.

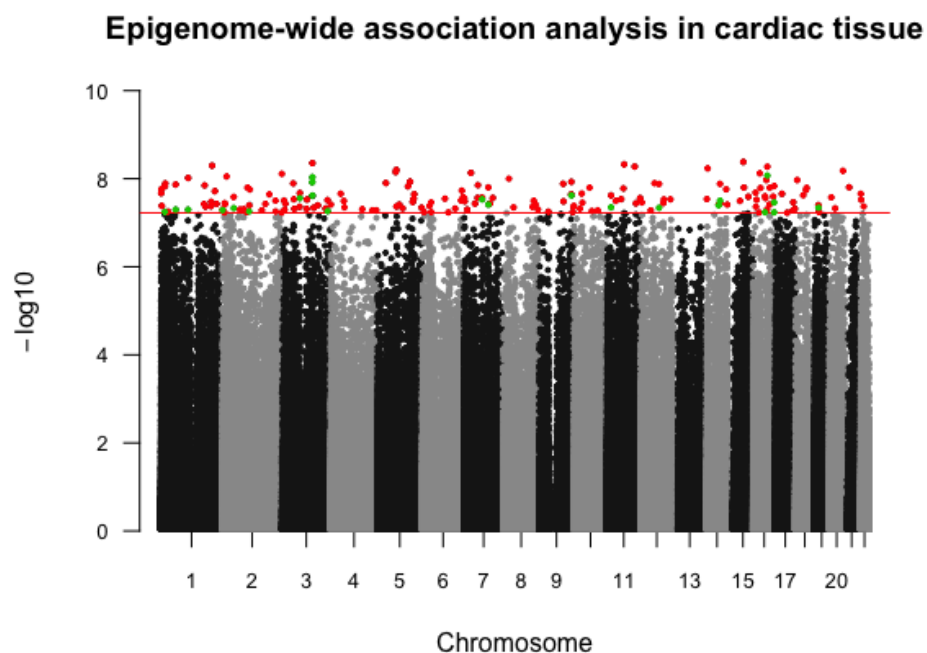


Figure 2: Manhattan plot for discovery-stage EWAS of DCM. The horizontal significance line (red) corresponds to the epigenome-wide significance threshold ($P < 5.96E-08$, $0.05/838,624$ tests). The 194 DCM sentinel CpGs are highlighted. At genomic loci with >1 epigenome-wide significant signal, secondary signals (green) were identified by conditioning on the lead signal (lowest P in region; red).

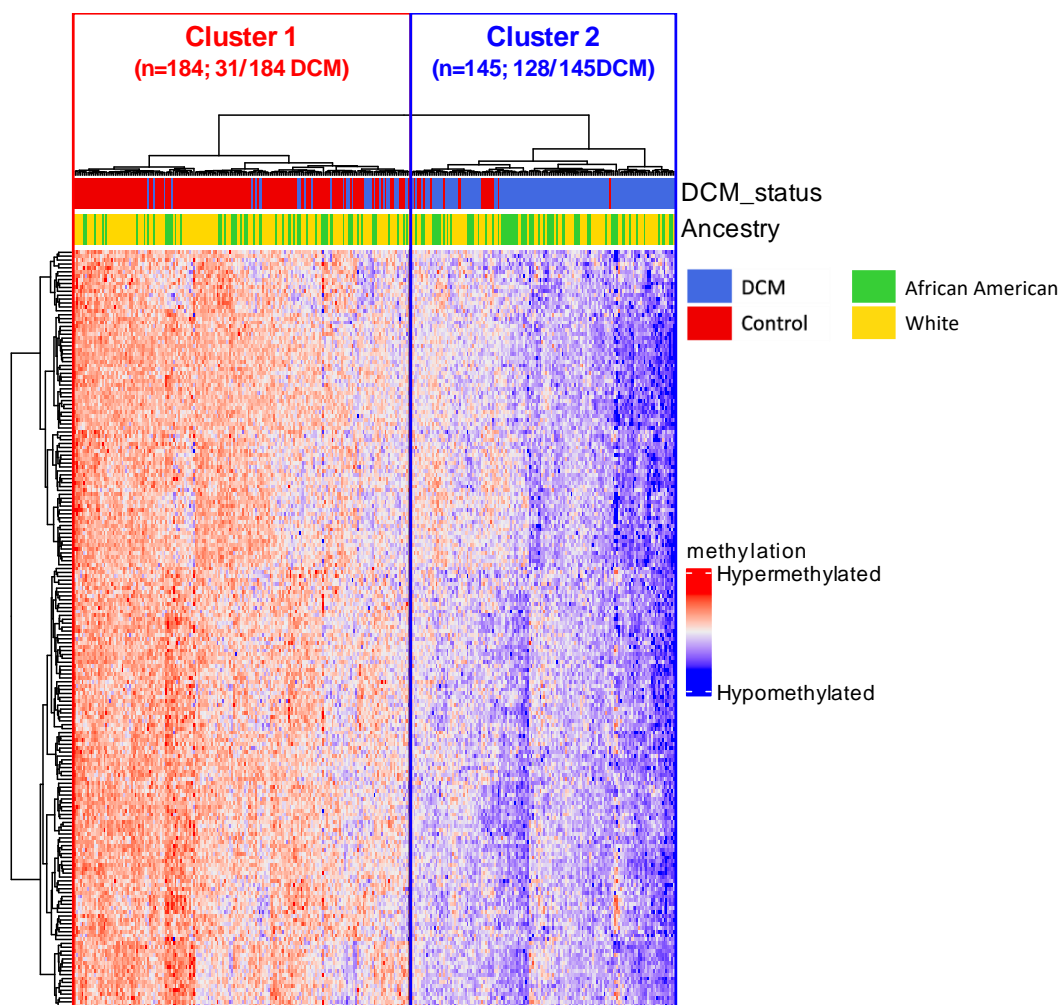


Figure 3: Unsupervised clustering using methylation levels of 194 sentinel CpGs. DCM, dilated cardiomyopathy.

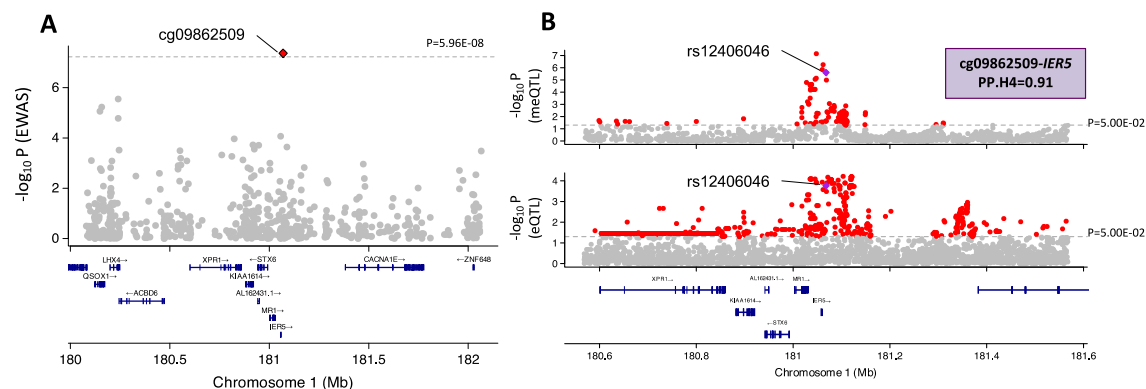


Figure 4 Regional plots for the sentinel CpG gene pair showing the highest posterior probability of colocalization. Two types of regional association plots are shown for *cg09862509-IER5*: **A. Associations of CpGs surrounding *cg09862509* with DCM from discovery-stage EWAS (MAGNet).** The displayed region is centered on *cg09862509* (red; bordered) **B. Genetic associations for *cg09862509* and *IER5*.** Genetic associations were obtained for a +/-500kb region centered on the instrumental variable used in SMR assess causal relationships between *cg09862509* methylation and *IER5* gene expression (meQTL from MAGNet; eQTL from GTEx.) Posterior probability of colocalization between *cg09862509* meQTLs and *IER5* eQTLs was 0.91. Gene tracks for all regional plots consist of protein-coding genes obtained from Ensembl Release 75 (last release based on hg19 assembly).

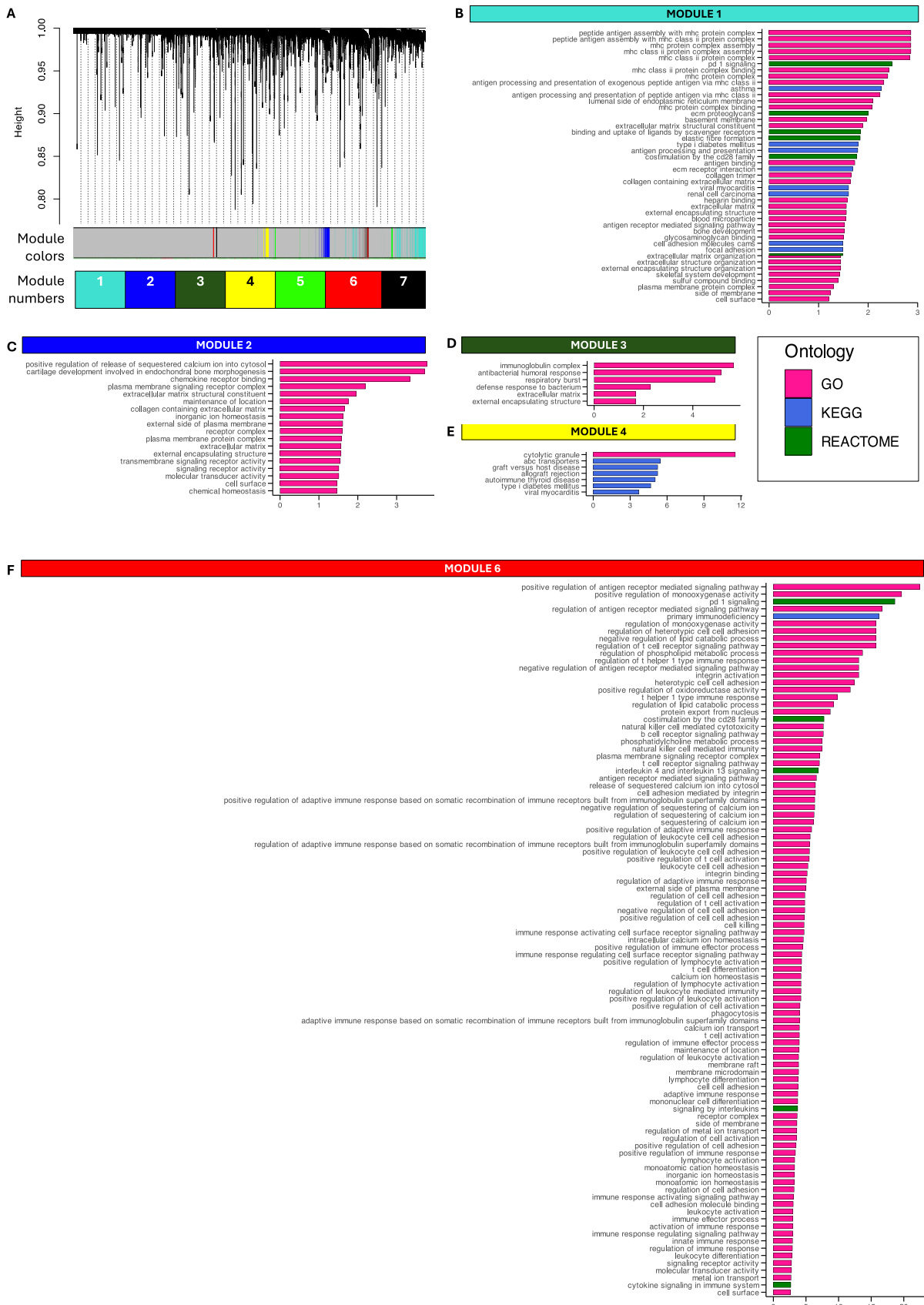


Figure 5. Co-methylation modules identified using DCM-associated CpGs. A. Hierarchical cluster tree (dendrogram) of co-methylation modules. WGCNA identified seven modules that exhibit conservation between co-methylation networks constructed independently within the White (n=209) and African American (n=118) ancestries. The color band underneath the tree indicates distinct modules (grey = CpG sites that are not clustered into any module). **B-F. Bar plots of over-represented gene sets (FDR $P < 0.05$) in co-methylation modules.** Genes were assigned to CpGs that were identified as replicated cis-eQTMs, signifying a strong correlation between methylation and gene expression (MAGNet FDR $P < 0.05$, consistent direction of association in BMCB analysis). Gene set enrichment within a module was assessed against a background set consisting of all genes mapped to CpG sites that were used to identify co-methylation patterns (9,321 unique genes). Enriched gene sets are displayed by decreasing order of fold-change (x-axis; calculated as the ratio of the proportion of overlapping genes linked to module-specific CpGs to the proportion of overlapping genes linked to background CpGs) GO = gene ontology; KEGG = Kyoto Encyclopedia of Genes and Genomes.

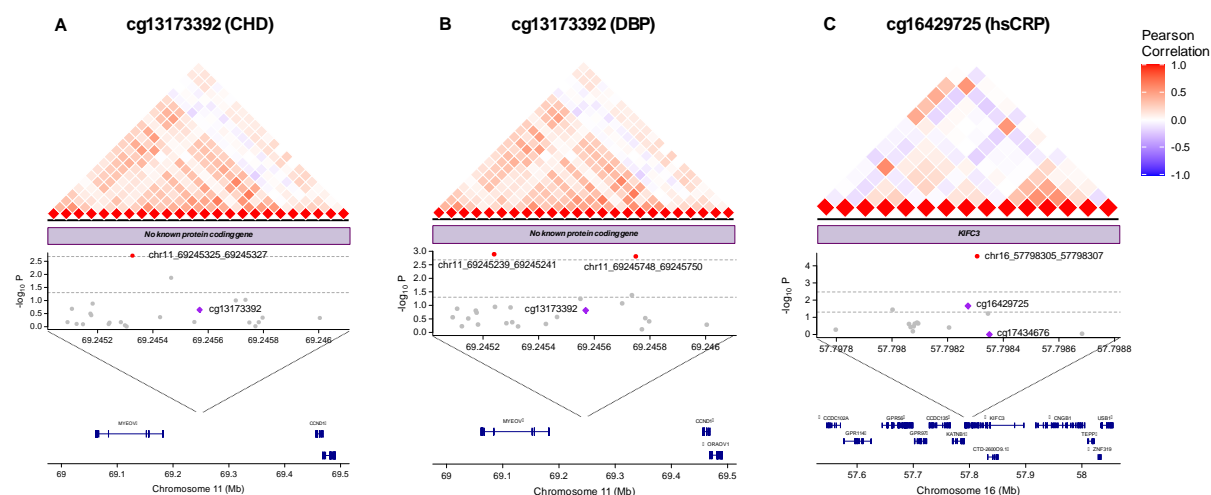


Figure 6. Regional associations of top DCM sentinel CpGs with CVD-related

traits. Regions surrounding DCM sentinel CpGs (+/-500bp) were fine-mapped to discover signals for CVD-related trait using blood samples from a population-based cohort (iHealth). To highlight the value of fine-mapping disease-associated loci, these plots illustrate the regional associations for cg13173392 and cg16429725, which are the sentinel CpGs with the strongest associations with DCM amongst those that underwent fine-mapping. Three types of plots are included in each panel:

(top) pairwise correlation in methylation levels between CpGs within a region;

(middle) regional associations between CVD traits and CpGs; (bottom) protein-

coding genes in the region. **A. Regional associations with coronary heart**

disease at the cg13173392 locus. This region consists of 24 CpGs. The lead signal

is chr11_69245325_69245327 ($P=1.93 \times 10^{-3}$). **B. Regional associations with**

diastolic blood pressure at the cg13173392 locus. This region consists of 24

CpGs. The lead signal is chr11_69245239_69245241 ($P=1.29 \times 10^{-3}$). A secondary

signal chr11_69245748_69245750 ($P=1.55 \times 10^{-3}$) was revealed after conditioning on

the lead signal. **C. Regional associations with the inflammatory mark high-**

sensitivity C-reactive protein at the cg16429725 locus. This region consists of 15

CpGs. The lead signal is chr16_57798305_57798307 ($P=2.79 \times 10^{-5}$). purple = DCM

sentinel CpGs; red = neighbouring CpGs associated with the respective CVD traits
(Bonferroni $P < 0.05$).